

Internet Robot Sistemi: Web tabanlı veriler, uygulamalar ve servisler için bir entegrasyon aracı

Taner KÖRPE

Siemens Business Services, Sistem Entegrasyonu Bölümü ODTÜ Teknokent ANKARA

taner.korpe@sbs.com.tr

Özet

Bu bildiri, web kaynaklarından veri toplama ve entegre etme işini yapan, SBS Türkiye’de geliştirilmiş bir uygulama olan Internet Robot Sistemi tanıtılacaktır. IRS, HTN (Hierarchical Task Network) planlama mimari yaklaşımını kullanmaktadır. Kullanıcı tarafından tanımlanan, hangi web kaynağına nasıl erişileceğini belirten plan dosyaları IRS Planner tarafından aynı anda ve özerk bir şekilde ifa edilir. IRS ile uygulama geliştirme olanaklarının kolaylaştırılabilmesi için, grafiksel kullanıcı arabirimi geliştirilmiştir. IRS, ticari bir ürün olarak kullanılmak üzere, birçok uygulama için configure edilmiştir.

Abstract

We introduce Internet Robot System (IRS), which is a web resource composer tool that has been developed in Siemens Business Services Turkey section. IRS users can develop valuable applications by composing various web resources such as data, applications and services. IRS uses hierarchical task network planning approach as the underlying formalism for the composition of web resources. User defined plan files, which describe what web resources are to be composed and how, are executed by the IRS planner in an autonomous and concurrent manner on the behalf of its users. Both to simplify the application development with IRS and to make the underlying composition formalism transparent to the users, a considerable effort has been spent on the development of visual interfaces of the tool. As a commercial tool, IRS has been configured and used in several different applications.

1.Giriş

Günümüz web ortamında, kullanıcıların ihtiyaçlarını karşılamak üzere web kaynaklarını kullanarak küçük ancak kullanışlı uygulamalar geliştirmeleri mümkündür. Web tabanlı uygulama geliştirmede web servis teknolojilerinin kullanımının en popüler metodlardan biri olmasına paralel olarak, web kaynaklarının entegrasyonu konusunda önceki çalışmalar da temel olarak web servisleri entegrasyonuna dayanmaktaydı. Sonuçta, semantik web servislerinin entegrasyonu için birçok yöntem ortaya çıktı [1]. Sektörden gelen tecrübemiz gösterdi ki, özellikle de e-devlet alanında, müşterilerimiz şu özelliklere sahip bir web entegrasyon aracına ihtiyaç duymaktadırlar: i) kullanıcılara web servislerinin haricinde HTML, HTTP tabanlı uygulamalar gibi veri kaynaklarını entegre etme yeteneği vermesi ii) geliştirme sürecini grafik arayüzlerle kolaylaştırma iii) web kaynakları entegrasyonu formalizasyonunun kullanıcılar tarafından anlaşılır olması. Bu özellik ayrıca, görsel arayüzlerle web

kaynağının tanımlanmasına olanak vererek desteklenecektir. Sanayi penceresinden baktığımızda, böyle bir aracın geliştirilmesi büyük öneme sahiptir. Siemens Business Services (SBS) Türkiye, üniversiteden akademik destek alarak, yukarıda özellikleri sıralanan bir web entegrasyon aracı geliştirmiştir. “Internet Robot System” (IRS) adı verilen araç, planlama modülü de içerdiğinden, otonom ve eş zamanlı olarak kullanıcının tanımladığı planları kullanıcı adına yerine getirme yeteneğine sahiptir.

Plan dosyaları ile hangi web kaynaklarının nasıl entegre edileceği tanımlanır. Bir web kaynağından diğer web kaynağına veri aktarımı, form gönderme işlemi ya da karşılıklı veri/parameter iletişimi kurulması bu dosyada yapılacak tanımlarla mümkündür. Böyle bir geliştirme aracını kullanırken şu temel adımlar takip edilmelidir. İlk adım birleştirilecek en küçük birimi temsil eden ilkel yapının tanımlanmasıdır. Bu özelleştirilmiş yapılara “görev” diyoruz. Buradaki yaklaşım, her derlenecek ve birleştirilecek web kaynağı için bir görev yapısı tanımlamak gerekliliği şeklindedir. İkinci adım, bu ilkel görev yapılarının nasıl karmaşık görevler olarak bir araya getirilebileceğinin tanımlanmasıdır. Karmaşık görevleri plan olarak adlandırabileceğimizden, hiyerarşik görev ağı (HTN), basit görevlerin karmaşık görevler olarak formalize edilmesi yöntemiyle tanımlanması mümkün olmuştur. Dolayısıyla IRS’de web kaynaklarının derlenmesinin formalizasyonunda HTN planlama yaklaşımı kullanılmıştır. Son adım ise, HTN tabanlı web kaynaklarının derlenmesini içeren karmaşık görev tanımlarını ifa edecek planlama modülünün geliştirilmesidir. Bildirinin geri kalan kısmında, IRS çekirdek mimarisini, IRS modüllerini ve mimaride belirtilen her katmanının açıklandığı mantıksal mimarisini inceleyeceğiz. Halihazırda birçok projede kullanılmış olan IRS’nin gerçek hayat kullanım örnekleri ve kullanım senaryolarından bahsedecek ve son olarak da sonuç ve planlanan çalışmalardan ve referanslardan söz edilecektir.

2.Çekirdek IRS Mimarisi

IRS web tabanlı verileri ve uygulamaları entegre edebilen, web kaynaklarından veri toplayabilen bir araçtır. IRS bu işleri yapmak için kullandığı görev tanımlarını ve planları otonom olarak ve aynı anda gerçekleştirebilir. Her görev tanımı, erişilecek web kaynağını ve alınacak verilecek verilerle ilgili detaylı bilgi içerirken plan tanımları da bu görevlerin birlikte nasıl ifa edileceğini belirten bilgileri içerir. Görev tanımları için DF (Task Definition Files) kullanılır. Bu dosyalar önceden tanımlanmış standart bir yapıdadır ve web kaynağı ile ilgili tüm bilgi ve fonksiyonluları içerir. IRS’de her tip görev için bir XML şema tanımlanır ve görev tanımı bu şemaya uygun olarak XML formatında DF olarak hazırlanır. Her görev dosyasının birbiriyle ilişkilendirilerek karmaşık görevler oluşturulabilmesi için formal bir model seçilmiştir. HTN (Hierarchical Task Network) [5] adı verilen, üst seviye karmaşık görevleri basit görevlerle ifade etme prensibine dayalı yöntem ile temel görevler palnlar haline getirilmiştir. Web kaynaklarını derlemek için kullanılan bu formal yöntemin haricinde, standartlara dayalı tanımlanmış bir sözdizim gereklidir. Bu amaçla görevlerin yerine getirilme sırasını, yapısını ve fonksiyonunu tanımlayan bir XML şema hazırlanmıştır. Bu şemaya uygun olarak hazırlanan planları (PF) yerine getirmek üzere IRS modüllerinden biri olan Planlama Modülü geliştirilmiştir.[3]

“WebTask” Görev Yapısı

Bu görev web’den veri erişimi, gözetme operasyonlarından ve web uygulamalarının çalıştırılmasından oluşur. Bu görevin çıktısı XML yapısında bir dokümandır. Bu görev için görev tanım şemasının kök

elementi “Connnection” elementidir. Bu element bağlanılacak olan web kaynağı ile ilgili temel özellikleri içerir ve birden fazla “Request” ve “Response” elementinden oluşur. Table-1’de “Connection” (bağlantı) elementinin özellikleri görülmektedir.

Özellik	Açıklama
URL	Web kaynağının URL adresi.
Name	Kullanıcı tarafından bağlantıya verilen mantıksal isim.
Port	Hedef bağlantının port numarası. Varsayılan port numarası 80, 443 ise SSL bağlantılar için ayrılmıştır.
Method	Bağlantı yöntemi (HTTP ya da HTTPS).
SaveFileName	Kayıt edilecek sonuç dosyasının adı
AcceptCookie	Çerez (Cookie) dosyaları aktif yada değil
ProxyHost	Varsa proxy yada firewall için URL/IP adresi
ProxyPort	Proxy yada Firewall için port numarası.
LogFileName	Log dosyasının adı
AuthUserName	Yetkilendirme için gerekli olan kullanıcı adı
AuthPassword	Yetkilendirme için gerekli olan şifre

Tablo 1. “Connection” (bağlantı) elementinin özellikleri

“Request” elementi web kaynağına gönderilecek istekle ilgili gerekli bilgiyi içerir. “Request” elementinin temel Table-2’de görülmektedir.

Özellik	Açıklama
Name	Kullanıcı tarafından Request’e verilen mantıksal isim.
Type	Request (istek) yöntemi (GET, POST, PUT)
IsAutonomous	Request’in periyodik olarak uygulanıp uygulanmamağını belirtir
TimePeriod	IsAutonomous değerine bağlı olarak, zaman periyodunun değerini taşır
IsActive	Request’in aktif olup olmadığı değerini taşır.

Tablo 2. “Request” (İstek) elementinin özellikleri

Request elementi “RequestParameter” elementini de barındırabilir. “RequestParameter” elementi bir web kaynağındaki HTML formdan otomatik olarak alınan yada kullanıcı tarafından sağlanan parameter değerlerini taşımak için kullanılır. Şekil-1’de belirlenen bir kaynaktan veri toplamak üzere hazırlanmış bir görev tanımını ifade eden “WebTask” XML dokümanı gösterilmektedir.

Şekil-1’deki XML dokümanında, firewall yada proxy sunucusu kullanılmadığı varsayılarak, “Connection” elementinin “ProxyPort” ve “ProxyHost” özellikleri kullanılmamıştır. Aynı şekilde hedef sitenin kullanıcı adı ve şifre istemeyen bir web kaynağı olduğu varsayıldığından “authorization” ve “authentication” özelliklerine de yer verilmemiştir.

```
<Connection xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="WebTask.xsd" URL="http://www.someURL" Name="SomeName" AcceptCookie="True" SaveFileName="c:\output\Results.xml" LogFileName="c:\output\Log.txt">
  <Request Name="getGoogle" Type="GET" IsAutonomous="true" IsActive="true" TimePeriod="20">
    <RequestParameter Name="par1" Value="val1"/>
    <RequestParameter Name="par2" Value="val2"/>
  </Request>
  :
</Connection>
```

Şekil 1. An example XML document, which shows the use of connection and request elements.

Sonuçta, “WebTask” görevinin çıktısını ifade etmek üzere “Connection” elementinin bir alt elementi olan “Response” elementi kullanılmıştır. “WebTask” görevinin çıktısı “Response” elementinin alt taglarına göre şekillenmiş olan bir XML dosyasıdır.

“Operation” Görev Yapısı

Bu görevin amacı web kaynağından elde edilen veriler üzerinde, dört işlem, mantıksal işlemlerle karşılaştırma, metni bölme ve parçalama gibi işlemler yapabilmek, sonuçları XML olarak kayıt etmektir. Bu görevi, diğer görevlerden veri alıp bunları işleme sokarak çıktıya veya diğer görevlere iletmek üzere hazırlayan aracı bir görev olarak kabul edebiliriz.

```
<Operation operator="OpCode">
  <input>
    <parameter>
      <name>.....</name>
      <type>.....</type>
    </parameter>
    :
  </input>
  <output>
    <!-- The same structure as in input tag to represent parameters -->
  </output>
</Operator>
```

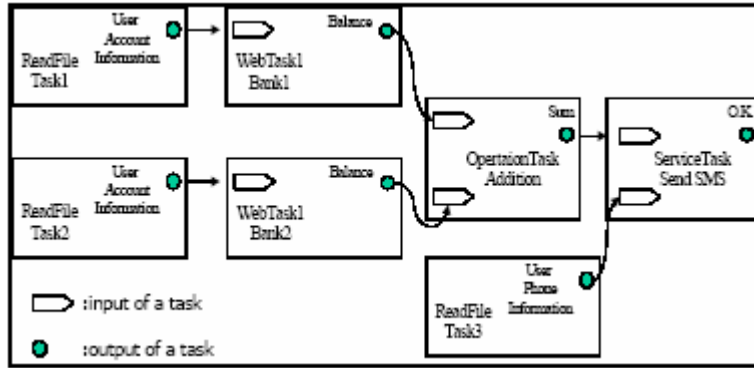
Şekil 3. “OperationTask” tanımlama yapısı

Şekil-3’de bu görevi tanımlayan örnek bir XML dokümanı görülmektedir.

“ServiceTask” adı verilen görev ile IRS web servislerine erişebilmektedir. Web servis istemleri bir arabirim ihtiyacı olmadan kullanıma olanak sağlayan DII (Dynamic Invocation Interface) kullanılarak gerçekleştirilmiştir[2].

Derleme Yalaşımı ve Planlama

Şu ana kadar tanımlanan web kaynaklarına erişimi sağlayan temel görev yapılarını inceledik. Bu bölümde birden fazla görev yapısının derlenerek karmaşık bir göreve yani plan haline gelmesi için kullanılan formalizasyon modeline değineceğiz. HTN (Hierarchical Task Network) planlama yaklaşımı [5] bu modelin temelini oluşturmaktadır. HTN yaklaşımı temelde basit görevleri kullanarak karmaşık görev (plan) tanımlamak prensibi üzerine kurulur. HTN yaklaşımı ayrıca web servislerinin de derlenmesi ve entegre edilmesinin planlanmasında kullanılır[4]. IRS kullanılarak geliştirdiğimiz yaklaşımda HTN sadece web servislerini değil, farklı türdeki web kaynaklarını da derlemek amaçlı kullanılmıştır. HTN yapısında görevler arasındaki ilişkiler bir görevin çıktısını diğer görevin girdisine bağlamak suretiyle tanımlanabilir. Basit bir HTN yapısı örneği Şekil-4’de verilmiştir.

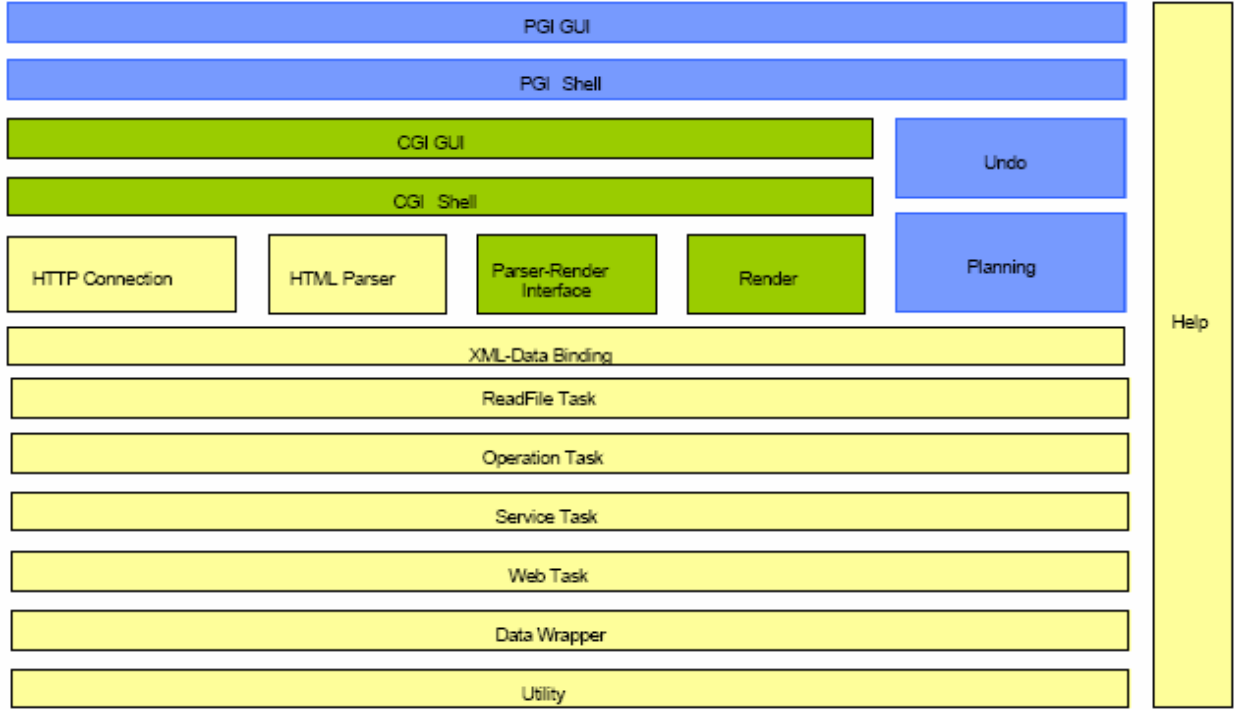


Şekil 4. Örnek bir HTN planı.

Şekil-4’de verilen yapıda, bir görev (WebTask) kullanıcının internet bankacılığı kaynağından aldığı hesap bilgisini “OperationTask” ile hesapta bir değişiklik olup olmadığını kontrol ederek başka bir göreve iletir. Bu görevde aldığı bilgiyi SMS sitesine bağlanarak kullanıcıya SMS atar. Banka sitesine ve SMS sitesine bağlanmak için gerekli olan hesap kullanıcı adı bilgileri Şekil-4’de “ReadTask” yardımı ile belirtilen XML dosyalarından okunur.

3.Mantıksal IRS Mimarisi

IRS’nin çekirdek mimarisinden sonra, IRS’nin genel tasarımı hakkında fikir verebilmek amacıyla genel mimariden ve modüllerinden bahsedeceğiz. Şekil-5’de ifade edilen yapı üç temel modülden oluşur. “Data Wrapper” modülü yerine getirilmiş görevlerin XML çıktılarını hazırlar. “Data Wrapper” modülünün üstündeki 4 modül temel görev yapılarını ifa etmek için gerekli kodu içerir. “XMLData Binding” modülü görev ve plan tanım dosyalarının XML olarak hazırlanmasından sorumludur.



Şekil 5. Genel IRS Yapısı ve modülleri

“Configuration Generation Interface- CGI” (DF oluşturma grafiksel arayüzü) kullanıcılara bir browser benzeri grafiksel arabirim ile görev tanım dosyası (DF) oluşturma olanağı verir. “Plan Generating Interface- PGI” (PF oluşturma grafiksel arayüzü) kullanıcılara grafik arabirim ile görevleri (DF dosyalarını) birbirlerini bağlama, görevler arasında operasyonel işlemler tanımlama ve bir plan dosyası oluşturma olanağı verir.

4.Uygulama Örnekleri

IRS birçok projede kullanılmış ticari bir üründür. Bu uygulamalardan bazıları hakkında bilgi verirken bazılarının ekran görüntülerini ve açıklamalarını sunacağız.

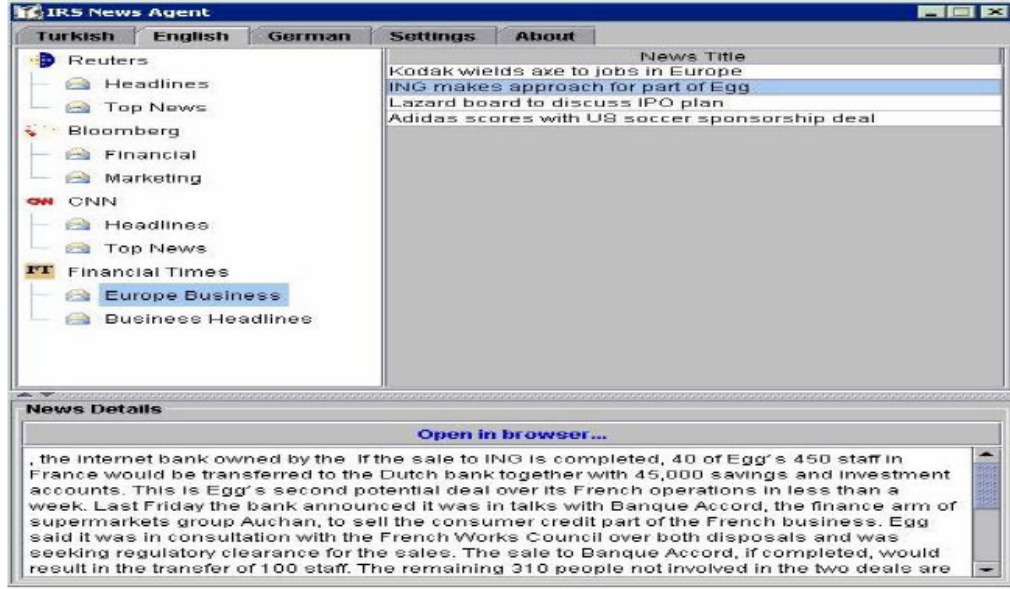
T.C. Kimlik Numarası Doğrulama

Bu uygulama ile web servisleri değil, TCKimlik numara sorgulaması yapılabilen web sitesi kullanılarak numara sorgulama ve verilen numaranın doğruluğunu control etme işini yapan bir IRS uygulaması geliştirilmiştir.

Haber Ajanı

Bu sistemde IRS, tanımlanan 3 farklı dilde, dünyanın en ünlü 10’den fazla haber sitesinden tanımlanan kategori ve başlıklarda haberleri XML olarak toplamakta ve kullanıcılara Şekil-6’da gösterilen arayüzle sunmaktadır. Ayrıca hangi zaman periyodunda haber kaynaklarının control edilmesi istendiği ve hangi

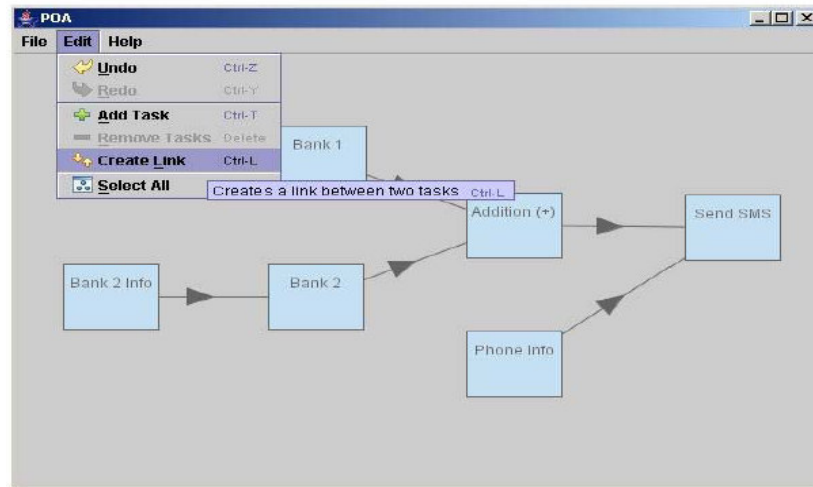
kaynaklardan veri toplanması istendiği gibi bilgileri kullanıcıların parametric olarak arayüz vasıtasıyla tanımlamasına olanak tanımıştır.



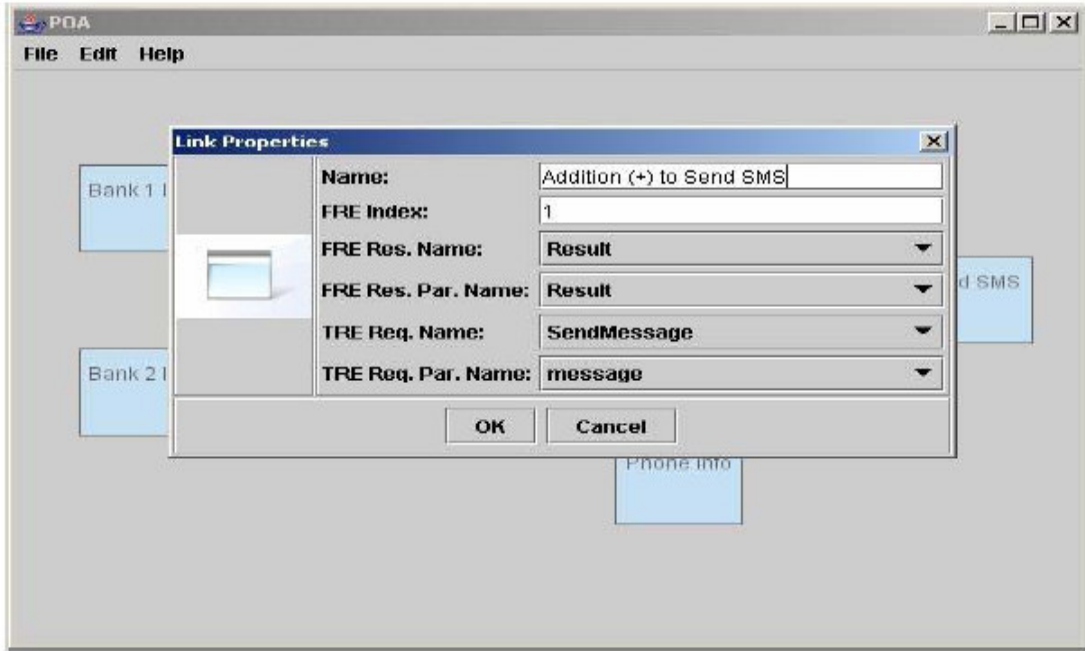
Şekil 6. Haber Ajansı ekran görüntüsü

Banka Hesap Yönetimi

Bu uygulama ile verilen banka hesabındaki değişimler IRS tarafından izlenerek bir değişim olduğunda kullanıcıya SMS atan bir uygulama geliştirilmiştir. IRS 15 saniyede bir verilen internet banka hesabına giriş yaparak control etmekte ve değişimleri SMS olarak bildirmektedir. IRS'nin bu uygulaması HTN yaklaşımı ile üretilen planın kullanıldığı bir örnektir. IRS'nin grafik arayüze sahip plan oluşturma arayüzü kullanılarak oluşturulan planın oluşturulma yöntemine ilişkin ekran görüntüsü Şekil-7'de, Şekil-8'de ve Şekil-9'da verilmiştir.

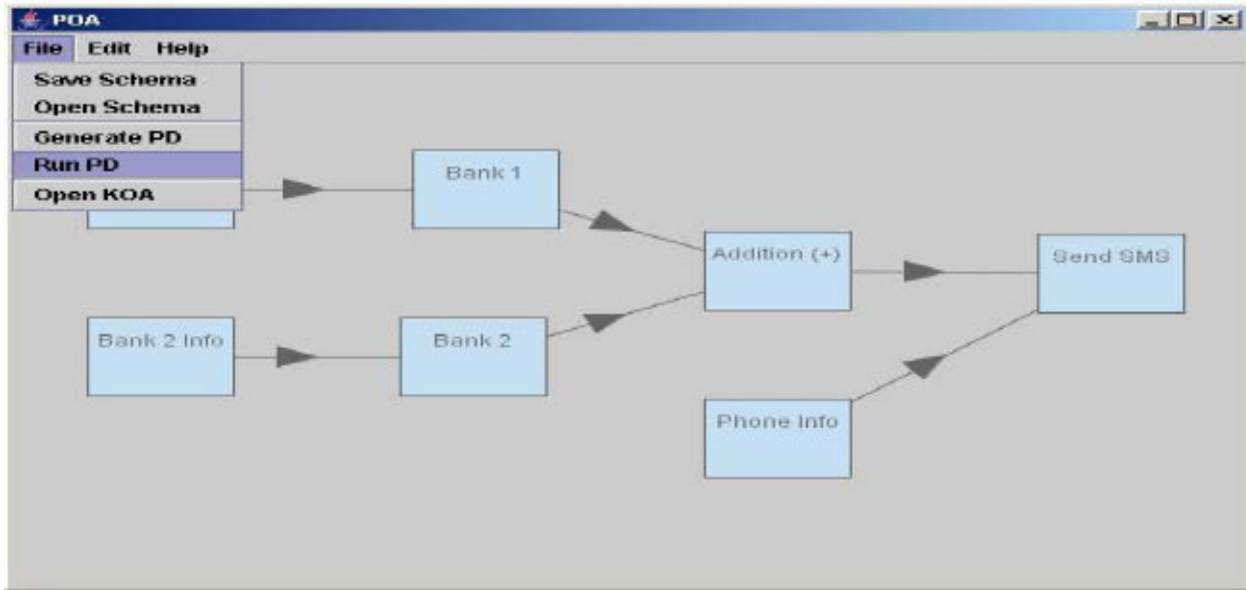


Şekil 7. Görevleri tanımlama ve ilişkilendirme arayüzü.



Şekil 8. İlişkinin özelliklerini belirleme arayüzü

Tüm tanımlamalar görsel arayüzlerle yapılmakta ve otomatik olarak XML olarak plan dosyası haline getirilmektedir. Oluşturulan plan Şekil-9’da görüldüğü gibi test edilmek üzere “Run PD” menüsü seçilerek çalıştırılabilmektedir. Bütün bu arayüzler IRS çekirdek modülünü kullanmak üzere Java ortamı kullanılarak implement edilmiş uygulamalardır.



Şekil 9. Planı oluşturma

5.Sonuç ve Planlanan Çalışmalar

Web sadece web servislerinden oluşmamaktadır. Web dünyasında birçok web uygulaması ve standartlardan bağımsız ciddi miktarda HTML verisi bulunmaktadır. Bu duruma ve müşteri geribildirimlerini değerlendirerek TÜBİTAK desteği ile geliştirmiş olduğumuz bu uygulama ile, kullanıcılar hertür web kaynağını entegre eden ve web kaynaklarından (HTML) XML formatında bilgi toplayabilen uygulamalar geliştirebileceklerdir. Gelecekte IRS'yi semantik web desteği olan bir araç haline getirmek üzere akademik araştırmalarımız ve çalışmalarımız devam etmektedir.

6. Kaynakça

- [1] Milanovic, N., ve Malek, M. Current solutions for web service composition. IEEE Internet Computing. 8(6):55-59, 2004.
- [2] Monson-Haefel, R. J2EE web services, Pearson Education, Inc., 2004.
- [3] Siemens Business Services. IRS planning module design document. SBS Turkey Application Development Bölümü, 2004.
- [4] Sirin, E, Parsia, B., Wu, D., Hendler, J., ve Nau, D. HTN planning for web service composition using SHOP2. Journal of Web Semantics, 1(4):377-396, 2004.
- [5] Williamson, M., Decker, K., ve Sycara, K. Unified information and control flow in hierarchical task networks. AAAI-96 notları ve "Theories of Action, Planning, and Control" konusunda çalıştay, Ağustos 1996. (http://www.ri.cmu.edu/pubs/pub_2178.html)